SparseFusion: Distilling View-Conditioned Diffusion for 3D Reconstruction

Year: 2022 | Citations: 252 | Authors: Zhizhuo Zhou, Shubham Tulsiani

---

Abstract

We propose SparseFusion, a sparse view 3D reconstruction approach that unifies recent advances in neural rendering and probabilistic image generation. Existing approaches typically build on neural rendering with reprojected features but fail to generate unseen regions or handle uncertainty under large viewpoint changes. Alternate methods treat this as a (probabilistic) 2D synthesis task, and while they can generate plausible 2D images, they do not infer a consistent underlying 3D. However, we find that this trade-off between 3D consistency and probabilistic image generation does not need to exist. In fact, we show that geometric consistency and generative inference can be complementary in a mode-seeking behavior. By distilling a 3D consistent scene representation from a view-conditioned latent diffusion model, we are able to recover a plausible 3D representation whose renderings are both accurate and realistic. We evaluate our approach across 51 categories in the CO3D dataset and show that it outperforms existing methods, in both distortion and perception metrics, for sparse-view novel view synthesis.